Systematic Review on Kinematic Assessments of Upper Limb Movements After Stroke

Schwarz, Anne ; Kanzler, Christoph M ; Lambercy, Olivier ; Luft, Andreas R ; Veerbeek, Janne M

Abstract: Background and Purpose- Assessing upper limb movements poststroke is crucial to monitor and understand sensorimotor recovery. Kinematic assessments are expected to enable a sensitive quantification of movement quality and distinguish between restitution and compensation. The nature and practice of these assessments are highly variable and used without knowledge of their clinimetric properties. This presents a challenge when interpreting and comparing results. The purpose of this review was to summarize the state of the art regarding kinematic upper limb assessments poststroke with respect to the assessment task, measurement system, and performance metrics with their clinimetric properties. Subsequently, we aimed to provide evidence-based recommendations for future applications of upper limb kinematics in stroke recovery research. Methods- A systematic search was conducted in PubMed, Embase, CINAHL, and IEEE Xplore. Studies investigating clinimetric properties of applied metrics were assessed for risk of bias using the Consensus-Based Standards for the Selection of Health Measurement Instruments checklist. The quality of evidence for metrics was determined according to the Grading of Recommendations Assessment, Development, and Evaluation approach. Results- A total of 225 studies (N=6197) using 151 different kinematic metrics were identified and allocated to 5 task and 3 measurement system groups. Thirty studies investigated clinimetrics of 62 metrics: reliability (n=8), measurement error (n=5), convergent validity (n=22), and responsiveness (n=2). The metrics task/movement time, number of movement onsets, number of movement ends, path length ratio, peak velocity, number of velocity peaks, trunk displacement, and shoulder flexion/extension received a sufficient evaluation for one clinimetric property. Conclusions- Studies on kinematic assessments of upper limb sensorimotor function are poorly standardized and rarely investigate clinimetrics in an unbiased manner. Based on the available evidence, recommendations on the assessment task, measurement system, and performance metrics were made with the goal to increase standardization. Further high-quality studies evaluating clinimetric properties are needed to validate kinematic assessments, with the long-term goal to elucidate upper limb sensorimotor recovery poststroke. Clinical Trial Registration- URL: https://www.crd.york.ac.uk/prospero/ . Unique identifier: CRD42017064279.

DOI: https://doi.org/10.1161/STROKEAHA.118.023531

Originally published at:
Schwarz, Anne; Kanzler, Christoph M; Lambercy, Olivier; Luft, Andreas R; Veerbeek, Janne M (2019). Systematic Review on Kinematic Assessments of Upper Limb Movements After Stroke. Stroke, 50(3):718-727.
DOI: https://doi.org/10.1161/STROKEAHA.118.023531

OPEN

Systematic Review on Kinematic Assessments of Upper Limb Movements After Stroke

Anne Schwarz, MSc*; Christoph M. Kanzler, MSc*; Olivier Lambercy, PhD; Andreas R. Luft, MD, PhD; Janne M. Veerbeek, PhD

Background and Purpose**—**Assessing upper limb movements poststroke is crucial to monitor and understand sensorimotor recovery. Kinematic assessments are expected to enable a sensitive quantification of movement quality and distinguish between restitution and compensation. The nature and practice of these assessments are highly variable and used without knowledge of their clinimetric properties. This presents a challenge when interpreting and comparing results. The purpose of this review was to summarize the state of the art regarding kinematic upper limb assessments poststroke with respect to the assessment task, measurement system, and performance metrics with their clinimetric properties. Subsequently, we aimed to provide evidence-based recommendations for future applications of upper limb kinematics in stroke recovery research.

Methods**—**A systematic search was conducted in PubMed, Embase, CINAHL, and IEEE Xplore. Studies investigating clinimetric properties of applied metrics were assessed for risk of bias using the Consensus-Based Standards for the Selection of Health Measurement Instruments checklist. The quality of evidence for metrics was determined according to the Grading of Recommendations Assessment, Development, and Evaluation approach.

Results**—**A total of 225 studies (N=6197) using 151 different kinematic metrics were identified and allocated to 5 task and 3 measurement system groups. Thirty studies investigated clinimetrics of 62 metrics: reliability (n=8), measurement error (n=5), convergent validity (n=22), and responsiveness (n=2). The metrics task/movement time, number of movement onsets, number of movement ends, path length ratio, peak velocity, number of velocity peaks, trunk displacement, and shoulder flexion/extension received a sufficient evaluation for one clinimetric property.

Conclusions**—**Studies on kinematic assessments of upper limb sensorimotor function are poorly standardized and rarely investigate clinimetrics in an unbiased manner. Based on the available evidence, recommendations on the assessment task, measurement system, and performance metrics were made with the goal to increase standardization. Further high-quality studies evaluating clinimetric properties are needed to validate kinematic assessments, with the long-term goal to elucidate upper limb sensorimotor recovery poststroke.

Clinical Trial Registration**—**URL: https://www.crd.york.ac.uk/prospero/. Unique identifier: CRD42017064279. **(*Stroke*. 2019;50:718-727. DOI: 10.1161/STROKEAHA.118.023531.)**

Key Words: biomechanical phenomena ■ movement ■ paresis ■ review ■ stroke ■ upper extremity

Deficits in upper limb sensorimotor function are experienced by about 80% of patients with stroke early after symptom onset.[1] Despite the availability of acute medical treatment and rehabilitation, upper limb impairment persists in about 60% of the patients 6 months poststroke.[2] These impairments can include muscle weakness, loss of interjoint coordination, and changes in muscle tone and sensation, which subsequently reduce the ability to use the upper limb when performing daily activities and increase dependency.[3,4] Understanding upper limb sensorimotor recovery poststroke is required to optimize therapy outcomes by developing effective interventions. One constraint impeding this understanding is the lack of standardized and responsive approaches to define and measure stroke-related upper limb deficits and their evolution.[5]

Traditionally, upper limb deficits poststroke are evaluated using established clinical assessments, such as the upper extremity subscale of the Fugl-Meyer Assessment (FMA-UE)[6,7] and the Action Research Arm Test.[8,9] A drawback of these assessments is that they are insufficiently sensitive to capture the quality of sensorimotor performance because of the use of

Received October 1, 2018; final revision received December 3, 2018; accepted January 11, 2019.

From the Division of Vascular Neurology and Neurorehabilitation, Department of Neurology, University Hospital Zurich, University of Zurich, Switzerland (A.S., A.R.L., J.M.V.); cereneo, Center for Neurology and Rehabilitation, Vitznau, Switzerland (A.S., A.R.L., J.M.V.); Biomedical Signals and Systems, Technical Medical Centre (TechMed Centre), University of Twente, Enschede, the Netherlands (A.S.); and Department of Health Sciences and Technology, Rehabilitation Engineering Laboratory, Institute of Robotics and Intelligent Systems, ETH Zurich, Switzerland (C.M.K., O.L.).

*A. Schwarz and C.M. Kanzler contributed equally.

The online-only Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.118.023531.

Correspondence to Anne Schwarz, MSc, Division of Vascular Neurology and Neurorehabilitation, Department of Neurology, University Hospital Zurich, University of Zurich, Frauenklinikstrasse 26, CH-8091 Zurich, Switzerland. Email anne.schwarz@usz.ch

ordinal scales. This impedes the ability to clearly distinguish behavioral restitution from compensation,[10,11] which is essential to understand neurological mechanisms of sensorimotor recovery poststroke. Behavioral restitution has been defined as a return toward more normal patterns of motor control with the impaired effector, whereas compensation strategies include new behavioral approaches by using intact muscles, joints, and effectors in the affected limb, to accomplish the desired task or goal.[12] Kinematic assessments promise to overcome these drawbacks by providing objective metrics that have the potential to sensitively capture movement quality and enable the monitoring of compensatory movements.[12–14] However, a variety of tasks, measurement systems, and kinematic metrics are used in clinical research. This limits comparability between studies and the potential for meta-analyses that are needed to establish a knowledge foundation about the mechanisms of upper limb recovery. Furthermore, information about clinimetric properties, such as reliability, measurement error, validity, and responsiveness of metrics derived from kinematic assessments is essential to confirm their physiological interpretation and robustness and thereby, their suitability for stroke recovery research.

Previous reviews summarized the use of kinematic metrics for the upper limb[15–20] and their physiological interpretation.[21] However, they focused only on specific measurement systems or did not differentiate metrics according to assessment tasks,[16,21] factors that are likely to influence the interpretation of kinematic metrics.[22] In addition, the majority of these reviews was not performed in a systematic way or did not rely on guidelines such as Preferred Reporting Items for Systematic Reviews and Meta-Analysis for reporting systematic reviews (PRISMA) and Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) for assessing risk of bias and grading the evidence.[23,24] Despite the importance of characterizing clinimetric properties, only 2 reviews investigated clinimetrics, but these focused solely on convergent validity between metrics and clinical scales[20] or did not consider assessment characteristics and the quality of the clinimetric evidence.[16]

This systematic review, therefore, aimed to provide a complete and unbiased overview of assessment tasks, measurement systems, and metrics with their clinimetric properties (reliability, measurement error, convergent validity, and responsiveness) for kinematic upper limb assessments poststroke. Subsequently, we proposed recommendations on how to design, evaluate, and apply kinematic assessments in future stroke recovery research.

Methods

This systematic review was registered in the international prospective register of systematic reviews PROSPERO (number CRD42017064279) and meets the Preferred Reporting Items for Systematic Reviews and Meta-Analysis requirements.[23] The search was performed in PubMed, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and Institute of Electrical and Electronics Engineers (IEEE) Xplore from inception to September 30, 2017. For the literature search in PubMed, see Table I in the online-only Data Supplement. The online-only Data Supplement contains detailed information on eligibility criteria, information sources, study selection, and data collection. The data that support the findings of this study are available from the authors on reasonable request.

Data Collection and Definitions

For each study, information about the kinematic assessment and clinimetric properties were extracted. Additionally, patient demographics, stroke-related information, and the level of upper limb impairment was recorded.

Assessment tasks were categorized into 5 groups based on the nature of the performed upper limb movements. Two-dimensional (2D) tasks in the horizontal plane were divided into 2D pointing (ie, discrete movements to defined targets) and 2D shape drawing (ie, continuous movements) tasks. Three-dimensional (3D) tasks were partitioned into 3D pointing and 3D reach-to-grasp (ie, discrete movements with object manipulation) tasks. Studies that could not be allocated to one of these groups were assigned to the other tasks group.

Measurement systems were categorized into 3 groups based on their expected influence on upper limb movements during the kinematic assessments. Influence refers especially to the interaction forces between measurement system and patient because of friction, inertia, and arm weight support. Group A contained measurement systems with minimal influence on movements, such as inertial measurement units and optical and electromagnetic motion capture systems used without arm weight support. Group B contained measurement systems expected to have medium influence, such as end effectors and motion capture systems used with arm weight support. Group C consists of measurement systems likely to have high influence, such as exoskeletons.[25]

Each reported kinematic metric (ie, a parameter extracted from kinematic data using specific postprocessing algorithms) was assigned to one of the following constructs based on their physiological interpretation: accuracy, data-driven scores, efficacy, efficiency, movement planning, precision, smoothness, spatial posture, speed, temporal posture, or workspace. Their definitions (see the online-only Data Supplement) were based on previous work,[16,21] descriptions in the included studies, and experience of the authors and were required to link metrics to their assumed physiological interpretation.

Study Quality Assessment

The risk of bias for studies investigating clinimetric properties of kinematic metrics was assessed using the COSMIN checklist for systematic reviews.[24] The clinimetric properties test-retest reliability (ie, proportion of measured variance that results from actual differences between patients), measurement error (ie, error not attributed to actual changes in the measured construct), convergent validity (ie, degree to which correlation of metrics to clinical scales is consistent with the hypothesis), and responsiveness (ie, ability to capture longitudinal changes in the measured construct) were analyzed.

Synthesis of Results

The results of the clinimetric evidence and study quality assessment were synthesized for each investigated metric across tasks by applying the Grading of Recommendations Assessment, Development and Evaluation principles.[24] Herewith, the evidence of multiple studies is summarized based on the risk of bias (ie, study quality), inconsistency (ie, contradicting results), and imprecision (ie, small population sizes). For reliability, intraclass correlation coefficients of ≥0.7 were considered to be sufficient[25] (ie, the evaluation of results was appropriate for this property). Measurement error was considered to be sufficient, if the smallest detectable change or limits of agreement were below the minimal important change. Convergent validity was evaluated analyzing correlation coefficients (r) between kinematic metrics and clinical scales. The FMA-UE was selected as reference clinical scale because it was most commonly reported for describing upper limb motor impairment (76% of the studies). For convergent validity, a moderate-to-very-high correlation ($|r| \geq 0.5$ with $P \leq 0.05$) between the FMA-UE and all metrics describing the physiological constructs accuracy, data-driven scores, efficacy, efficiency, smoothness, spatial posture, speed, temporal posture, and workspace led to a sufficient evaluation. For metrics describing another physiological construct, convergent validity could not be analyzed because it would require different reference scales that were typically not reported. For responsiveness, an area under the curve of ≥0.7 was sufficient. The evidence per clinimetric property per kinematic metric was evaluated according to the COSMIN criteria for good measurement properties (sufficient, insufficient, inconsistent, or indeterminate).[24] Outcomes were the summarized evidence (sufficient, indeterminate, or insufficient) and the quality of evidence (high, moderate, low, or very low) per kinematic metric and clinimetric property. Metrics were recommended for future use if the quality of the evidence was at least moderate and the summarized evidence was sufficient.

Results

Kinematic Upper Limb Assessments

The literature search resulted in 225 included studies (N=6197; Figure 1; Table II in the online-only Data Supplement). The included studies, as well as the participant and kinematic assessment characteristics, are listed in the online-only Data Supplement. According to our task classification, 81 studies used a 2D pointing task, 16 a 2D shape drawing task, 67 a 3D pointing task, 50 a 3D reach-to-grasp task, and 24 a task belonging to the other tasks group (Figure I in the online-only Data Supplement). Kinematic recordings were made with a measurement system of group A, B, and C in 130, 69, and 26 studies, respectively. In total, 151 different kinematic metrics (Figures 2 and 3; Figures II and III, and Table II in the online-only Data Supplement) were reported to quantify upper limb sensorimotor function. Figures 2 and 3 provide an overview of the frequency distribution of each kinematic metric per task, the assigned physiological construct, and the reported clinimetric properties.

2D Pointing Tasks

Patients (N=2536) included in studies using 2D pointing tasks had a median FMA-UE score of 34.35 (interquartile range [IQR], 22.40–47.59; reported in n=57). Eighty-two different kinematic metrics were used, all of them describing trunk, shoulder, and elbow movements (Figure 2A). The 5 most commonly assessed physiological constructs were smoothness (n=95), speed (n=78), efficiency (n=68), movement planning (n=60), and accuracy (n=48). The 5 most commonly used metrics were peak velocity (n=35), task/movement time (n=31), mean velocity (n=28), number of velocity peaks (n=21), and end point error (n=20).

2D Shape Drawing Tasks

Patients (N=817) included in studies reporting 2D shape drawing tasks had a median FMA-UE score of 33.40 (IQR, 22.00–45.69; reported in n=13). Thirty-two different kinematic metrics were reported, all of them describing trunk, shoulder, and elbow movements (Figure 2B). The 5 most commonly assessed physiological constructs were smoothness (n=18), accuracy (n=12), precision (n=12), speed (n=11), and efficiency (n=5). The 5 most commonly used metrics were mean velocity (n=8), trajectory error (n=6), axes ratio (n=5), normalized mean velocity (n=4), and normalized jerk (n=4).

3D Pointing Tasks

Patients (N=1818) included in 3D pointing tasks had a median FMA-UE score of 43.53 (IQR, 37.38–48.35; reported in n=48). Forty-nine different kinematic metrics were presented, all of them describing trunk, shoulder, and elbow movements. The 5 most commonly assessed physiological constructs were spatial posture (n=136), efficiency (n=85), speed (n=50), smoothness (n=32), and movement planning (n=27). The 5 most commonly used metrics were task/movement time (n=43), peak velocity (n=35), elbow flexion/extension angle (n=33), shoulder flexion/extension angle (n=31), and path length ratio (n=26).

3D Reach-to-Grasp Tasks

Patients (N=1178) performing a 3D reach-to-grasp task had a mean FMA-UE score of 46.00 (IQR, 37.40–52.35; reported in n=32). Sixty-six different kinematic metrics were reported. Forty-three metrics described trunk, shoulder, and elbow movements and 23, wrist, hand, and finger movements. The 5 most commonly assessed physiological constructs were spatial posture (n=79), efficiency (n=59), grasping efficiency (n=39), speed (n=34), and smoothness (n=27). The 5 most commonly used metrics were task/movement time (n=38), peak velocity (n=29), peak grip aperture (n=23), elbow flexion/extension angle (n=19), and time to peak velocity (n=19).

Other Tasks

Patients (N=593) involved in other task assessments had a mean FMA-UE score of 27.35 (IQR, 24.40–39.23; reported in n=6). Forty-two different metrics were reported (Figure IV in the online-only Data Supplement). Thirty-eight metrics described trunk, shoulder, and elbow movements and 5, wrist, hand, and finger movements. The 5 most commonly assessed physiological constructs were spatial posture (n=25), spatial posture of hand, wrist, and finger (n=14), efficiency (n=11), accuracy (n=9), and smoothness (n=9). The 5 most commonly used metrics were trajectory error (n=7), task/movement time (n=6), wrist flexion/extension angle (n=6), elbow flexion/extension angle (n=5), and success rate (n=4).

Risk of Bias Assessment

The results of the risk of bias assessment can be found in Table IV in the online-only Data Supplement.

Synthesis of Evidence for Clinimetric Properties

Thirty (13.3%) studies investigated ≥1 clinimetric properties of 62 (41.1%) kinematic metrics. In total, 124 (20.5%) of 604 possible combinations of all metrics and clinimetric properties were evaluated. Table 1 displays the metrics/clinimetric properties with at least moderate quality of evidence and (in)sufficient summarized evidence (details and all results in Table V in the online-only Data Supplement).

Test-Retest Reliability

Test-retest reliability was analyzed for 30 (19.9%) kinematic metrics. The summarized evidence was sufficient for 21, indeterminate for 2, and insufficient for 7 metrics. The quality of evidence was moderate for 1, low for 8, and very low for 21 metrics. The only metric with a sufficient summarized evidence and of at least moderate quality was peak velocity.

Measurement Error

Measurement error was evaluated for 27 (17.9%) kinematic metrics. The summarized evidence was indeterminate for all metrics. The quality of evidence was moderate for 4, low for 10, and very low for 13 metrics.

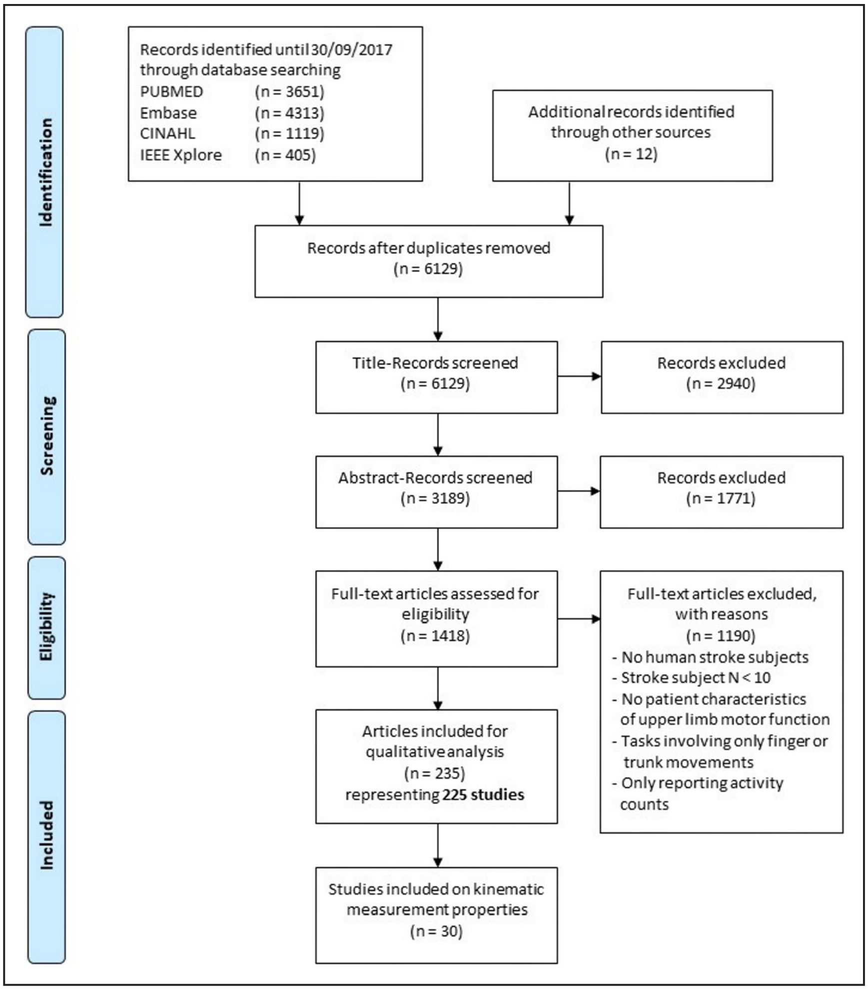

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flowchart, the systematic literature search. Adapted from Moher et al[23] with permission.

Convergent Validity

Convergent validity with the FMA-UE was analyzed for 58 (38.4%) metrics. The summarized evidence was sufficient for 22, indeterminate for 34, and insufficient for 2 metrics. The quality of evidence was high for 3, moderate for 11, low for 17, and very low for 27 metrics. Metrics with a sufficient summarized evidence and of at least moderate quality were the number of movement onsets/ends, task/movement time, path length ratio, number of velocity peaks, shoulder flexion/extension angle, and trunk displacement. Range of velocity was the only metric with insufficient summarized evidence and moderate quality.

Responsiveness

Responsiveness was evaluated for 9 (6.0%) metrics. The summarized evidence for responsiveness was sufficient for 3 and indeterminate for 6 metrics. The quality of evidence was very low for all metrics.

Discussion

This systematic review aimed to summarize the usage of tasks, measurement systems, and metrics for upper limb kinematic assessment poststroke, as well as the available evidence on the clinimetric properties of these metrics. We identified 225 studies, which we assigned to 5 task types and 3 measurement system groups. One hundred fifty-one kinematic metrics covering different aspects of upper limb sensorimotor function were reported. However, their clinimetric properties were only investigated in 30 studies, leading to mostly very low or low quality of evidence. Most of these studies investigated convergent validity (38.4% of the metrics) and reliability (19.9% of the metrics). These findings demonstrate the need for better standardization and evaluation of kinematic assessments.

There are several possible reasons for this missing standardization and clinimetric evidence. First, researchers tend to focus on the development of novel metrics rather than trying to use and validate existing ones. This may partly result from the scarce reporting on data processing methods and the dependency of some metrics on specific hardware that is not widely available. Second, systematically investigating clinimetric properties requires carefully designed studies and involves large numbers of subjects and resources,[24] which can be challenging to provide in practice. For example, the Grading of Recommendations Assessment, Development and Evaluation approach of COSMIN requires evidence from at least 100 patients across studies per metric and clinimetric property to avoid downgrading the quality of evidence.[24]

Previous reviews on upper limb kinematics have not led to specific guidelines for kinematic assessments poststroke.[15–21] With the aim of improving standardization, we defined evidence-based recommendations for designing and reporting kinematic upper limb assessments for stroke recovery research (Table 2). These should enhance comparability

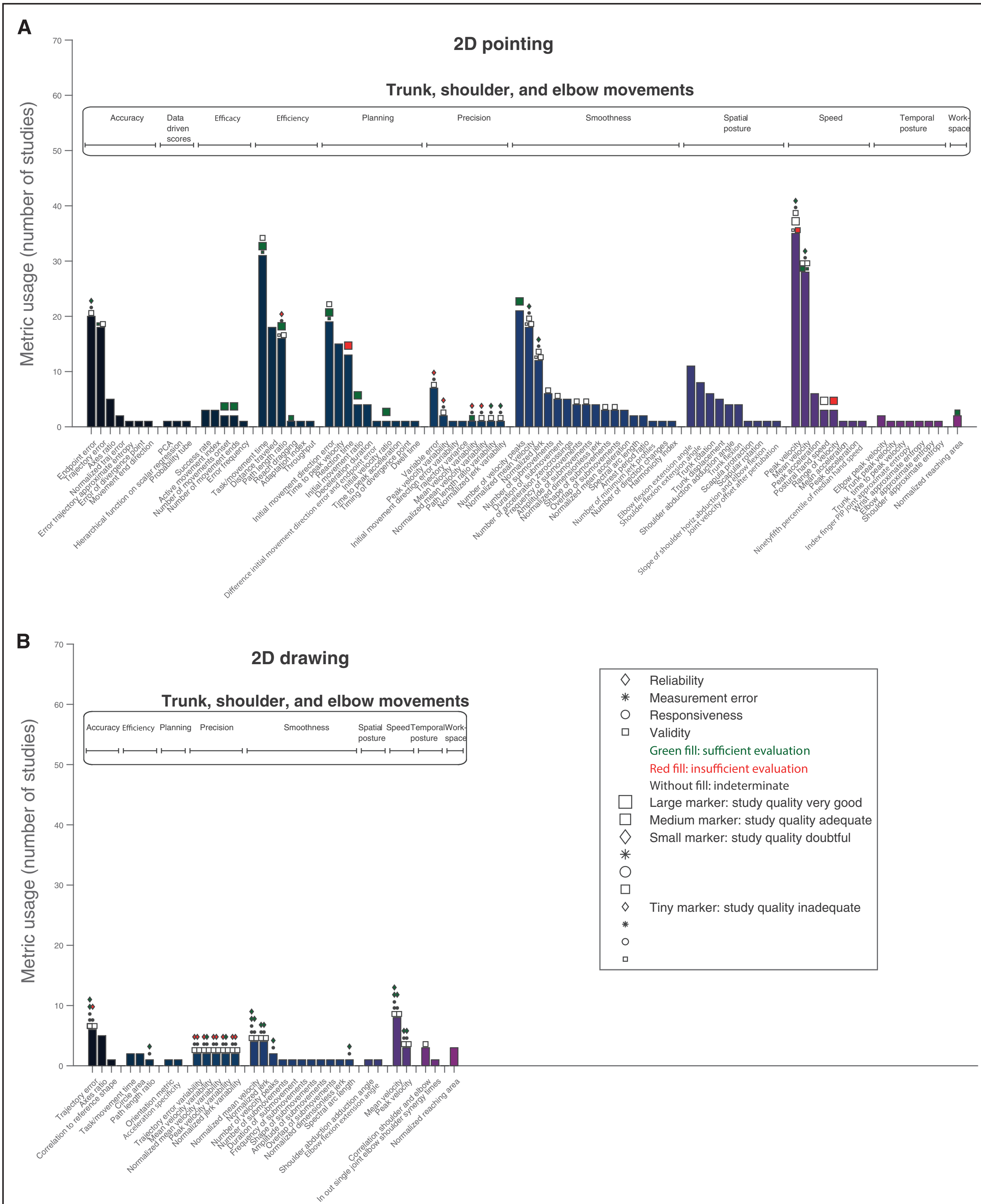

Figure 2. Kinematic metrics and clinimetric properties for the tasks 2D pointing (**A**) and 2D drawing (**B**). Metrics were grouped according to their assumed physiological interpretation. Type, size, and fill color of the annotated symbols indicate the evaluated clinimetric property (reliability, measurement error, responsiveness, or validity), study quality (inadequate, doubtful, adequate, or very good), and evaluation results (negative, indeterminate, or positive) of single studies, respectively.

between studies in the future and enable statistically summarizing study results in meta-analyses. We further advocate that researchers, clinicians, and funding agencies put more effort and resources in studies focusing on the evaluation of clinimetric properties of existing kinematic metrics. These actions should enable the research community to better exploit the

Table 1. Overview of the Kinematic Metrics and Their Clinimetric Properties

Kinematic Metric	Clinimetric Property	Quality of Evidence	Summarized Evidence	Quantitative Evidence
No. of movement onsets	Validity (+)	Moderate	Sufficient	|r|: −0.54
No. of movement ends	Validity (+)	Moderate	Sufficient	|r|: −0.58
Task/movement time	Validity (+)	High	Sufficient	|r|: −0.60; −0.60; −0.53; −0.52
Path length ratio	Validity (+)	Moderate	Sufficient	|r|: −0.54; 0.85
No. of velocity peaks	Validity (+)	Moderate	Sufficient	|r|: −0.58
Shoulder flexion/extension angle	Validity (+)	Moderate	Sufficient	|r|: 0.50; 0.56; 0.59; 0.70
Trunk displacement	Validity (+)	Moderate	Sufficient	|r|: −0.76; −0.72; −0.68
Range of velocity	Validity (–)	Moderate	Insufficient	|r|; −0.4
Peak velocity	Reliability (+)	Moderate	Sufficient	ICC: 0.74; 0.74; 0.87; 0.87; 0.93; 0.93; 0.94; 0.95; 0.95

Metrics/properties are shown for which the quality of evidence (ie, quality of the available studies) was at least moderate and the summarized evidence (ie, quality of the clinimetric evaluation results) was either sufficient (+) or insufficient (–). References can be found in the online-only Data Supplement. ICC indicates intraclass correlation coefficient; and |r|, correlation coefficient.

potential of kinematic assessments, which should help to provide objective measures characterizing components of poststroke recovery and fine-grained metrics that could better evaluate physiological changes because of rehabilitation interventions.

Application of Kinematic Upper Limb Assessments

When planning to apply upper limb kinematic assessments, it is essential to clearly define research questions and hypotheses that should guide the choice of task type, measurement system, and metrics. For example, to assess possible compensatory movement strategies, a 3D movement task requiring the coordination of shoulder abduction, elbow extension, and hand opening and closing (eg, 3D reach-to-grasp) should be favored because these movements are known to elicit pathological synergy patterns.[5,12,13] Compensatory movements can then be quantitatively captured by kinematic parameters describing the spatial posture of trunk, shoulder, elbow, and hand, as well as the efficiency of movements. Hence, a set of metrics including trunk displacement, shoulder flexion/extension angle, shoulder abduction/adduction angle, elbow flexion/extension angle, wrist flexion/extension angle, and path length ratio might be appropriate to capture these movements patterns,[12] as reflected by their frequent application in 3D tasks (Figure 3). This is also supported by the moderate-to-high correlation coefficients between most of these metrics and the FMA-UE—a clinical measure of pathological joint coupling, reported in this (Table 2; Table V in the online-only Data Supplement) and previous reviews.[16,20] Furthermore, the selected assessment task should be self-contained. Tasks trained during therapy, such as robot-assisted therapy, should be avoided because they confound results about upper limb function by including task-specific learning effects.[26] Data collected within therapy sessions should be exclusively used for monitoring performance during therapy and automatically adapting the difficulty level.[27] Additionally, the measurement system should have limited influence on the performed movements to best capture patient behavior during the task. For example, the inertia and arm weight support of an active exoskeleton could influence the synergistic coupling of the shoulder, elbow, and hand, thereby affecting the validity of kinematic outcomes.[17,28,29]

For assessing quality of sensorimotor performance, recommendations are proposed in Table 2 to guide the choice of metrics based on the available clinimetric evidence, the frequency of use, and insights from motor control, technical, and clinical perspectives. The clinimetric evidence could only rarely be used as a single criterion for the recommendations because of the mostly very low and low quality of the evidence. This further underlines the need for systematically evaluating clinimetric properties. Hence, it is of high priority that kinematic metrics that are often reported in the literature (Figures II through IV in the online-only Data Supplement), but poorly described in terms of clinimetric properties, are evaluated. This knowledge could help to better establish kinematic upper limb assessments poststroke.

Lastly, we want to emphasize the importance of reporting and discussing the influence of task type, measurement system, and task context (ie, therapy task or self-contained assessment task) on the results because these factors can challenge the comparison of metrics across studies. We further recommend that researchers report definitions of kinematic metrics, including equations, targeted physiological construct, signal processing methods, and clinimetric evidence, to foster transparency and thereby standardization.

Evaluation of Clinimetric Properties

For the first time, a methodological approach was applied for systematically investigating clinimetric properties of kinematic upper limb assessments. This allowed identifying misconceptions in study design and execution for most scientific publications, which led to low study quality according to COSMIN standards.[24] Hence, we recommend that researchers design and report clinimetric studies according to standardized guidelines like COSMIN. Nevertheless, the evaluation of some clinimetric properties remains challenging. For example, the comparison between clinical assessments and kinematic measures for analyzing convergent validity requires the choice of clinical assessments that

Table 2. Recommendations for Kinematic Upper Limb Assessments Poststroke

Assessment task	Should be hypothesis driven—task follows a research question
	Should correspond to the physical capabilities of the patient population
	Some voluntary upper limb movement but FMA-UE <30: 2D task
	FMA-UE ≥30: 3D task; finger flexion/extension required for 3D reach-to-grasp task
	Should, in case of an intervention trial, not coincide with the therapy task
Measurement system	Should have minimal influence on upper limb movements
Kinematic metrics	Should be hypothesis driven—metrics selection follows a research question and corresponds to physiological constructs of interest
	Should be selected based on available clinimetric evidence, the frequency of use, and insights from motor control, technical, and clinical perspectives
	Trunk/shoulder/elbow movements
	Accuracy: trajectory error, end point error
	Efficacy: number of movement onsets,* number of movement ends,* success rate
	Efficiency: task/movement time,* path length ratio,* distance traveled
	Planning: time to peak velocity, reaction time, initial movement direction error
	Precision: variable error
	Smoothness: number of velocity peaks,* normalized dimensionless jerk, spectral arc length
	Spatial posture: trunk displacement,* shoulder flexion/extension angle,* shoulder abduction/adduction angle, elbow flexion/extension angle
	Speed: peak velocity*
	Temporal posture: elbow peak velocity, time to peak elbow extension angle, correlation shoulder and elbow, trunk movement time, trunk peak velocity
	Workspace: normalized reaching area
	Wrist/hand/finger movements
	Accuracy†
	Efficacy†
	Efficiency: peak grip aperture, aperture path ratio, grasp time, grasp release time
	Planning: time to peak grip aperture
	Precision†
	Smoothness: normalized dimensionless jerk grasp aperture
	Spatial posture: wrist flexion/extension angle, maximal vertical wrist position, wrist adduction/abduction angle, finger extension angle
	Speed: peak velocity of grasp aperture
Reporting	Research questions, hypotheses, patient population, task, measurement system, kinematic metrics, and positioning/instructions of the subject should be described; implementation of metrics should be transparent (equations and processing steps)
Methodology	Evaluating clinimetric properties is urgently needed and should be performed according to standardized guidelines (eg, COSMIN)

2D indicates 2 dimensional; 3D, 3 dimensional; COSMIN, Consensus-Based Standards for the Selection of Health Measurement Instruments; and FMA-UE, Fugl-Meyer Assessment Upper Extremity subscale.

*Metrics with at least moderate quality of evidence and sufficient summarized evidence in one clinimetric property.

†No recommendation could be made.

capture the content of the physiological construct described by the kinematic metrics. This can be complex because of the often unclear relationships between clinical and kinematic assessments and the high amount of resources required to apply a battery of these assessments.[30] It, nevertheless, seems to be inadequate to expect very high correlations between kinematic metrics and clinical scales. Kinematic assessments are assumed to provide sensitive and objective readouts without ceiling effects,[31] whereas clinical scales are mostly of ordinal nature with low resolution and often have ceiling effects.[7] We can expect that kinematic assessments provide complementary information to clinical scales, which might lead to lower convergent validity. The evaluation of other aspects of validity, such as the comparison between

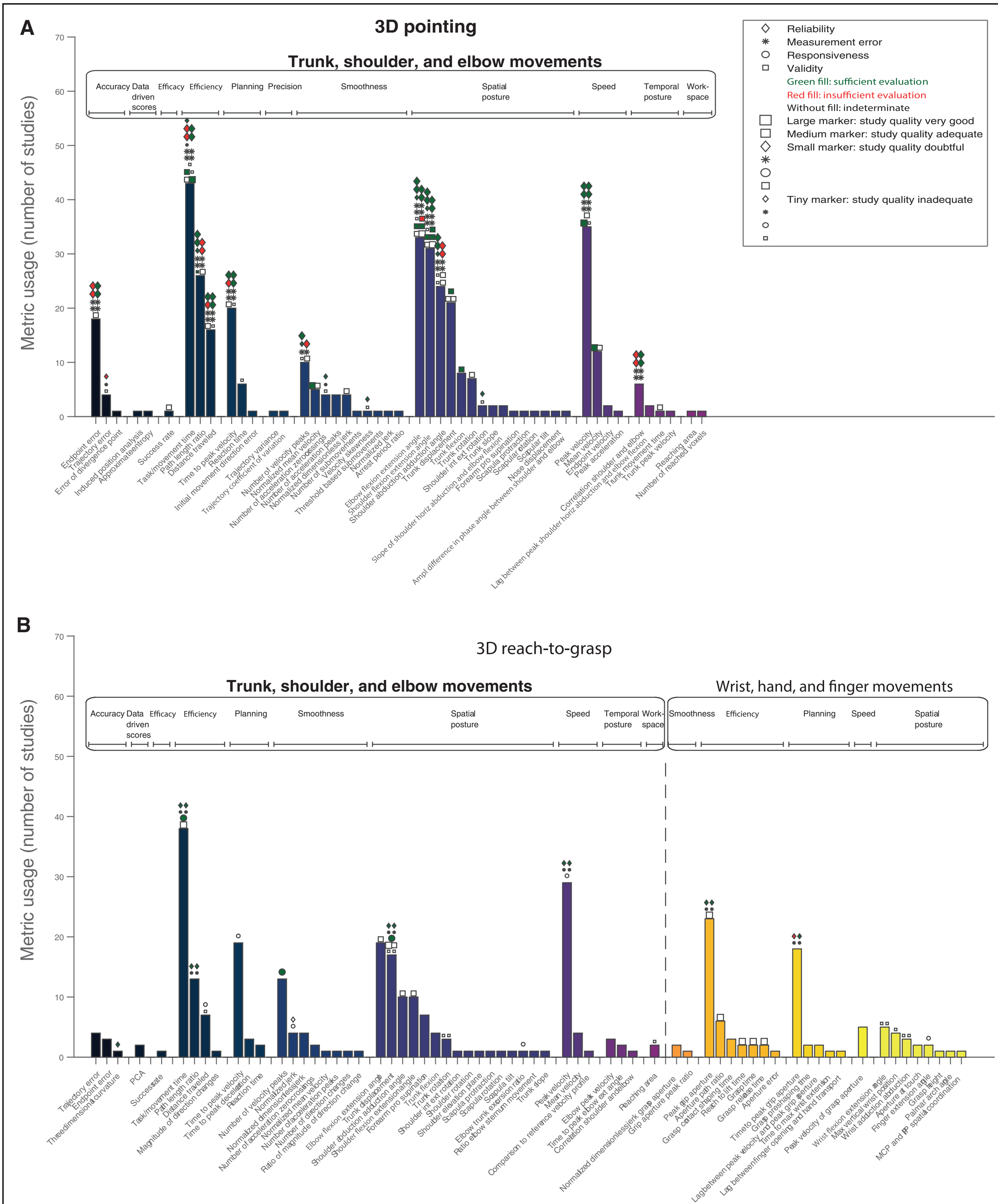

Figure 3. Kinematic metrics and clinimetric properties for the tasks 3D pointing (**A**) and 3D reach-to-grasp (**B**). Metrics were grouped according to the analyzed body part and their assumed physiological interpretation. Type, size, and fill color of the annotated symbols indicate the evaluated clinimetric property (reliability, measurement error, responsiveness, or validity), study quality (inadequate, doubtful, adequate, or very good), and evaluation results (negative, indeterminate, or positive) of single studies, respectively.

patients with stroke and healthy age-matched controls (ie, discriminative or known group validity[24]), should, therefore, also be considered.[16]

Acknowledging the relevance of reliability and responsiveness when investigating physiologically relevant changes during recovery, it is of utmost importance that more effort will be put into increasing evidence-based evaluations of these clinimetric properties.

Limitations

We suggested a classification for assessment tasks based on the nature of the performed movements, for measurement systems based on their influence on movements, and for metrics based on the assumed physiological interpretation. We acknowledge that this classification could have been implemented differently, although our suggestions were based on the literature, descriptions provided by the studies, and experience of the authors. In addition, the convergent validity analysis was conducted solely by comparing kinematic metrics and the FMA-UE because other clinical assessments were not consistently reported.

Conclusions

Although upper limb kinematic assessments are frequently used in stroke research, there is a lack of standardization for the use of assessment tasks, measurement systems, and kinematic metrics, as well as a paucity of high-quality studies analyzing clinimetrics. We underlined important considerations and proposed recommendations for designing and reporting of kinematic assessments after stroke, as well as for performing studies to evaluate clinimetric properties. These recommendations aim to enhance standardized and evidence-based kinematic upper limb assessments, with the long-term goal to elucidate upper limb recovery poststroke.

Acknowledgments

We thank Dr Werner L. Popp for the insightful discussions and Christopher Jarrett for language editing.

Sources of Funding

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 688857 (SoftPro), from the Swiss State Secretariat for Education, Research and Innovation (contract number 15.0283-1), and the P&K Pühringer Foundation. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Disclosures

None.

References

1. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review. *Lancet Neurol*. 2009;8:741–754. doi: 10.1016/S1474-4422(09)70150-4
2. Nijland RH, van Wegen EE, Harmeling-Van der Wel BC, Kwakkel G; EPOS Investigators. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS cohort study. *Stroke*. 2010;41:745–750. doi: 10.1161/STROKEAHA.109.572065
3. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. *Lancet*. 2011;377:1693–1702. doi: 10.1016/S0140-6736(11)60325-5
4. Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke*. 2011;42:1482–1488. doi: 10.1161/STROKEAHA.110.604090
5. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Neurorehabil Neural Repair*. 2017;31:784–792. doi: 10.1177/1545968317732662
6. Fugl-Meyer AR, Jääskö L, Norlin V. The post-stroke hemiplegic patient. II. Incidence, mortality, and vocational return in Göteborg, Sweden with a review of the literature. *Scand J Rehabil Med*. 1975;7:73–83.
7. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer Assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair*. 2002;16:232–240. doi: 10.1177/154596802401105171
8. Carroll D. A quantitative test of upper extremity function. *J Chronic Dis*. 1965;18:479–491.
9. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: properties of the Action Research Arm Test. *Arch Phys Med Rehabil*. 2006;87:1605–1610. doi: 10.1016/j.apmr.2006.09.003
10. Chen HM, Chen CC, Hsueh IP, Huang SL, Hsieh CL. Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. *Neurorehabil Neural Repair*. 2009;23:435–440. doi: 10.1177/1545968308331146
11. Lin KC, Chuang LL, Wu CY, Hsieh YW, Chang WY. Responsiveness and validity of three dexterous function measures in stroke rehabilitation. *J Rehabil Res Dev*. 2010;47:563–571. doi: 10.1682/JRRD.2009.09.0155
12. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: the Stroke Recovery and Rehabilitation Roundtable taskforce. *Int J Stroke*. 2017;12:444–450. doi: 10.1177/1747493017711816
13. Krebs HI, Krams M, Agrafiotis DK, DiBernardo A, Chavez JC, Littman GS, et al. Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. *Stroke*. 2014;45:200–204. doi: 10.1161/STROKEAHA.113.002296
14. Lambercy O, Maggioni S, Luenenburger L, Gassert R, Bolliger M. Robots and wearable sensor technologies for measurement/clinical assessment. In: *Neurorehabilitation Technology*. 2nd ed. Berlin, Germany: Springer Verlag; 2016:183–207. doi: 10.1007/978-3-319-28603-7
15. de los Reyes-Guzmán A, Dimbwadyo-Terrer I, Trincado-Alonso F, Monasterio-Huelin F, Torricelli D, Gil-Agudo A. Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: a review. *Clin Biomech (Bristol, Avon)*. 2014;29:719–727. doi: 10.1016/j.clinbiomech.2014.06.013
16. Alt Murphy M, Häger CK. Kinematic analysis of the upper extremity after stroke – how far have we reached and what have we grasped? *Phys Ther Rev*. 2015;20:137–155. doi: 10.1179/1743288X15Y.0000000002
17. Ellis MD, Lan Y, Yao J, Dewald JP. Robotic quantification of upper extremity loss of independent joint control or flexion synergy in individuals with hemiparetic stroke: a review of paradigms addressing the effects of shoulder abduction loading. *J Neuroeng Rehabil*. 2016;13:95. doi: 10.1186/s12984-016-0203-0
18. Shishov N, Melzer I, Bar-Haim S. Parameters and measures in assessment of motor learning in neurorehabilitation; a systematic review of the literature. *Front Hum Neurosci*. 2017;11:82. doi: 10.3389/fnhum.2017.00082
19. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. *J Neuroeng Rehabil*. 2017;14:20. doi: 10.1186/s12984-017-0229-y
20. Tran VD, Dario P, Mazzoleni S. Kinematic measures for upper limb robot-assisted therapy following stroke and correlations with clinical outcome measures: a review. *Med Eng Phys*. 2018;53:13–31. doi: 10.1016/j.medengphy.2017.12.005
21. Nordin N, Xie SQ, Wünsche B. Assessment of movement quality in robot- assisted upper limb rehabilitation after stroke: a review. *J Neuroeng Rehabil*. 2014;11:137. doi: 10.1186/1743-0003-11-137
22. Subramanian SK, Yamanaka J, Chilingaryan G, Levin MF. Validity of movement pattern kinematics as measures of arm motor impairment poststroke. *Stroke*. 2010;41:2303–2308. doi: 10.1161/STROKEAHA.110.593368
23. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62:1006–1012. doi: 10.1016/j.jclinepi.2009.06.005
24. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27:1171–1179. doi: 10.1007/s11136-017-1765-4

25. Just F, Özen Ö, Bösch P, Brobovsky H, Klamroth-Marganska V, Riener R, et al. Exoskeleton transparency: feed-forward compensation vs. disturbance observer. *Automatisierungstechnik*. 2018;66:1014–1026. doi: 10.1515/auto-2018-0069
26. Schweighofer N, Wang C, Mottet D, Laffont I, Bakhti K, Reinkensmeyer DJ, et al. Dissociating motor learning from recovery in exoskeleton training post-stroke. *J Neuroeng Rehabil*. 2018;15:89. doi: 10.1186/s12984-018-0428-1
27. Metzger JC, Lambercy O, Gassert R. Performance comparison of interaction control strategies on a hand rehabilitation robot. *IEEE Int Conf Rehabil Rob*. 2015:846–851.
28. Lan Y, Yao J, Dewald JPA. The impact of shoulder abduction loading on volitional hand opening and grasping in chronic hemiparetic stroke. *Neurorehabil Neural Repair*. 2017;31:521–529. doi: 10.1177/1545968317697033
29. Kanzler CM, Martinez Gomez S, Rinderknecht MD, Gassert R, Lambercy O. Influence of arm weight support on a robotic assessment of upper limb function. IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob 2018). 2018:1–6. doi: 10.1109/BIOROB.2018.8487682
30. McKenzie A, Dodakian L, See J, Le V, Quinlan EB, Bridgford C, et al. Validity of robot-based assessments of upper extremity function. *Arch Phys Med Rehabil*. 2017;98:1969.e2–1976.e2. doi: 10.1016/j.apmr.2017.02.033
31. Thrane G, Alt Murphy M, Sunnerhagen KS. Recovery of kinematic arm function in well-performing people with subacute stroke: a longitudinal cohort study. *J Neuroeng Rehabil*. 2018;15:67. doi: 10.1186/s12984-018-0409-4